

✓
高等学校教材

信息论基础

傅祖芸 编



電子工業出版社

信 息 论 基 础

傅 祖 芸 编

電 子 工 業 出 版 社

内 容 提 要

本书系高等学校工科电子类专业基础课统编教材。全书分七章，系统地讲述香农信息论的基本原理。内容包括：信息的基本概念，信息理论的形成与发展，熵的概念，信道与信道容量，平均失真度与信息率失真函数，以及信源编码和信道编码的基本原理与定理。

本书文字通俗，概念清晰，既可作为高等学校电子类信息论课程的教材，也可供计算机、人工智能、生物医学工程、管理科学等专业研究生及有关工程技术人员参考。

信 息 论 基 础

傅祖芸 编

责任编辑 哈利

电子工业出版社出版（北京市万寿路）

新华书店北京发行所发行 各地新华书店经售

山东电子工业印刷厂印刷

开本：787×1092 1/16 印张：12.625 字数：291千字

1986年4月第1版 1986年5月第1次印刷

印数：10000册 定价：2.15元

统一书号：15290·308

出 版 说 明

根据国务院关于高等学校教材工作分工的规定，我部承担了全国高等学校工科电子类专业课教材的编审、出版的组织工作。从一九七七年底到一九八二年初，由于各有关院校，特别是参与编审工作的广大教师的努力和有关出版社的紧密配合，共编审出版了教材159种。

为了使工科电子类专业教材能更好地适应社会主义现代化建设培养人才的需要，反映国内外电子科学技术水平，达到“打好基础、精选内容、逐步更新、利于教学”的要求，在总结第一轮教材编审出版工作经验的基础上，电子工业部于一九八二年先后成立了高等学校《无线电技术与信息系统》、《电磁场与微波技术》、《电子材料与固体器件》、《电子物理与器件》、《电子机械》、《计算机与自动控制》、中等专业学校《电子类专业》、《电子机械类专业》共八个教材编审委员会，作为教材工作方面的一个经常性的业务指导机构。并制定了一九八二到一九八五年教材编审出版规划，列入规划的教材、教学参考书、实验指导书等共217种选题。在努力提高教材质量，适当增加教材品种的思想指导下，这一批教材的编审工作由编审委员会直接组织进行。

这一批教材的书稿，主要是从通过教学实践、师生反映较好的讲义中评选择优和从第一轮较好的教材中修编产生出来的。广大编审者，各编审委员会和有关出版社都为保证和提高教材质量作出了努力。

这一批教材，分别由电子工业出版社、国防工业出版社、上海科学技术出版社、西北电讯工程学院出版社、湖南科学技术出版社、江苏科学技术出版社、黑龙江科学技术出版社和天津科学技术出版社承担出版工作。

限于水平和经验，这一批教材的编审出版工作肯定还会有许多缺点和不足之处，希望使用教材的单位、广大教师和同学积极提出批评建议，共同为提高工科电子类专业教材的质量而努力。

电子工业部教材办公室

前 言

本教材是编者根据多年的教学实践编写的，曾作过多次修改，现经无线电技术与信息系统教材编审委员会无线电通信编审小组评选审定，并推荐出版。

本教材由中国科技大学研究生院傅祖芸编写，编审小组委托华中工学院蔡德钧负责主审。编审者均依据编审小组审定的修编大纲进行修编和审阅。

随着科学技术的发展，知识体系不断更新，人类的知识结构也正在发生着变化。由于信息概念的深化，信息理论愈来愈被人们所重视。信息论作为一门专业基础课，已不局限于通信专业，而且也是无线电技术、自动控制、计算机、以及生物医学工程、遗传工程、人工智能、心理学、语言学、管理科学等专业必修课之一。根据这一指导思想，教材着重阐述香农理论的基本概念、基本分析方法及其主要结论。除必要的数学描述及论证外，尽量减少了繁琐的数学推导，结合例题，由浅入深地引出了基本概念与原理。为了配合教学，还搜集了一定数量的习题附在各章后面。

教材的参考教学时数为40~60学时。全书共分七章。在各章中，标有“*”号的小节均属于加宽加深的内容，可根据不同的教学要求和学时数，灵活讲授。

在教材编写过程中，曾得到陈宗骞教授的热情指导和具体帮助，对本书内容提出了许多宝贵意见。本书的绪论部分中也较多引用了陈教授1979年在中国科大研究生院上课的讲义。同时还得到了汪漱玉副教授和龙文澄教授的热情指导和帮助。在教材审编会议上，陆大铨、吴伯修、张宏基、葛果行、樊昌信、宋文涛等编委对教材大纲和内容提出了宝贵的意见。主审人对全书进行了认真细致的审阅。在编辑过程中还得到厉森梁同志的帮助。编者在此一并致谢。

由于水平有限，书中难免存在错误和不妥之处，恳请广大读者批评指正。

编 者

1985·9

目 录

前 言

第一章 绪 论

- 1.1 信息的概念 1
- 1.2 信息论研究的对象、目的和内容 5
- 1.3 信息论发展简史和现状 7

第二章 离散信源

- 2.1 信源的数学模型及分类10
- 2.2 离散信源的信息熵12
- 2.3 熵的基本性质17
- 2.4 离散无记忆的扩展信源19
- 2.5 离散平稳信源22
- 2.6 马尔可夫信源30
- 2.7 信源的相关性和剩余度34
- 2.8 信息熵与热力学熵, 意义信息和加权熵37
- 习题.....39

第三章 离散信道

- 3.1 信道的数学模型和分类43
- 3.2 信道疑义度及平均互信息47
- 3.3 平均互信息的特性49
- 3.4 离散无记忆的扩展信道54
- 3.5 串联信道的互信息60
- 3.6 信道容量65
- 3.7 信源与信道的匹配77
- 习题.....78

第四章 无失真信源编码

- 4.1 编码器82
- 4.2 等长码和等长信源编码定理83
- 4.3 变长码89
- 4.4 变长信源编码定理96
- 4.5 变长码的编码方法103
- 习题109

第五章 有噪信道编码

- 5.1 错误概率和译码规则.....114
- 5.2 错误概率与编码方法.....118

5.3 有噪信道编码定理.....	125
习题.....	129
第六章 保真度准则下的信源编码	
6.1 引言.....	131
6.2 信息率失真函数的定义和性质.....	132
6.3 信息率失真函数的计算.....	140
6.4 保真度准则下的信源编码定理.....	144
习题	154
第七章 连续信源和连续信道	
7.1 连续信源的统计特性和离散化.....	156
7.2 连续信源的熵.....	158
7.3 具有最大熵的连续信源.....	163
7.4 熵功率.....	166
7.5 连续信源熵的变换.....	166
7.6 噪声.....	170
7.7 连续信道的信息传输率和信道容量.....	172
7.8 连续信源的信息率失真函数.....	177
习题	181
附录A 凸函数和詹森不等式	183
附录B 马尔可夫链	187
附录C 熵函数计算用的函数表	194
参考文献	196

第一章 绪 论

信息论是在长期的通信工程的实践中,与概率论、随机过程和数理统计这些数学学科相结合而逐步发展起来的一门新兴科学。随着信息概念的不断深化,它在科学技术上的重要性早已超越了狭义的通信工程的范畴,在许多领域中日益受到科学工作者的重视。

本章首先引出信息的概念,进而讨论这一学科的研究对象、目的和内容,并简述本学科的历史发展、现状和动向。

1.1 信息的概念

在信息论中,信息是最基本、最重要的概念。

科学来源于实践,科学上许多术语往往来自日常生活,信息这个术语就是如此,它是在人类社会互通情报的实践过程中产生的。到了本世纪四十年代末,随着信息论这一学科的诞生,信息这一名词就成了科学上的概念,在含意上它比日常生活中所说的信息更为深刻了。

在日常生活中,信息是指“消息”、“情况”、“知识”、“情报”等等。例如,当人们收到一封电报、接到一个电话,或早晨从收音机里听到气象预报,或看了电视里的新闻之后,人们就说得到了“信息”。但到底得到了多少信息呢?对于这个问题,有时仅仅有一些模糊的似乎可以比较的概念而无确切的回答。例如,当打电话的对方告诉我的消息是我早就知道了的,我就觉得这个消息对我没有什么意义,或者说没有给我什么信息。相反,如果电话传来的消息,不但是我原来所不知道的,而且完全出乎我的意料,就会觉得这个信息给我的印象很深,给了我很多信息。又如,甲乙两人同去听某一学者讲课,由于两人的业务基础不同,他们听到的虽然是同一内容,但听后所得到的新知识,或者说所得到的有用的信息往往是不一样的。这些例子说明,在日常生活中,信息这个用语缺乏确切的概念,而且有很强的主观性。科学上所说的信息正是从这个原始的、含糊不清的概念中加以概括、提炼、提高和开拓而得到的。它有严格、确切的含义,有一定的数学模式,并能定量地量度。

在日常生活中,我们收到一封电报、接到一个电话或听到气象预报后,所得到的具体内容是对某一事物状态的描述。例如,气象预报说:“晴间多云”,这是对气象状态这一事物的具体描述。电报上的报文:“母病愈”,这是对母亲身体健康情况的描述。这些都是人脑以外的客观物质世界表现出来的各种不同的运动状态。而电话中说:“他想去上海”,这是存在于他头脑里的思维活动,反映了人的主观世界——大脑物质的思维运动所表现出来的思维状态。

以文字、语音、图象等这些能够为人们的感觉器官所感知的物理现象,把客观物质运动和主观思维活动的状态表达出来就成为消息。从通信的观点出发,构成消息的各种物理现象要具有两个条件:一是能为通信双方所理解,二是可以传递。因此从电报、电

话或电视中得到的是一些具体的消息，是描述主、客观各种事物运动状态的消息。而电报、电话、电视等则都是这些消息的传递系统。

现在，我们进一步来考察在这些消息传递过程中的一些共同特点和内在规律。

任何消息传递系统(电报、电话、传真、电视、雷达、视觉神经、听觉神经等)都有一个发出消息的发送端(一般称为信源)和接收消息的接收端(一般称为收信者或信宿)以及连接两者的通道(一般称为信道)，如图1.1所示。



图 1.1 消息传递系统

对收信者来话，是一个从不知到知的过程，或者说是一个从不确定到确定的过程。如果没有这样一个特点，那就根本不需要通信系统了。试想，如果收信者在收到电报或电话之前就已经知道报文的内容或电话的内容，那还要电报系统、电话系统干什么呢？

由于消息是对客观和主观的各种事物运动状态的具体描述，所以从通信过程来看，收信者的所谓“不知”就是不知道发送端将发送描述何种运动状态的消息。例如打电报，报文“母病愈”是母亲身体健康状态的一种描述，而母亲身体健康情况会表现出不同的状态。可见收信者在看到报文以前存在一种“不确定性”，他不能确定母亲身体健康状态如何。一旦看到报文以后，只要报文是清楚的，在传递过程中没有差错，那么他原来的“不确定性”就没有了，他就获得了信息。所以通信过程是一种从不确定到确定的过程，是消除不确定性的过程。不确定性消除了，收信者就获得了信息。由此可见，信息是事物运动状态(或它的存在方式)的不确定性的描述。因此，在消息传递系统中实质上传输的是信息，而消息只是表达信息的工具。

这里就有一个定量地测度信息的问题。显然，信息量与不确定性的程度有关。

用数学的语言来讲，不确定性就是随机性，具有不确定性的现象就是随机现象。因此，可通过研究随机现象的数学工具——概率论与随机过程来测度不确定性大小。

例如，如果发报人甲事前已经与收报人乙商定，为了报告某件事是否发生，甲只要在报文中告诉乙“是”与“否”即可。换句话说发报人甲只要在2个汉字“是”与“否”中选择一个即可。如果选择“是”与“否”的机会是均等的，即概率各为 $1/2$ ，在这种情况下，收报人在接到报文以前的不确定性是不大的，只不过是“二中择一”。从这个例子看出，消息中的不确定性的与可能发生的消息的数目及各消息发生的概率有关。

再如，气象预报，我们知道可能出现的气象状态有许多种。以十月份北京地区为例，最多的是“晴间多云”，其它则有“晴”、“多云”、“多云转阴”、“阴”、“阴有小雪”，但一般说来，“下小雪”的概率是极小的，“下大雪”的可能性则更小。因此，在听气象预报前，我们一般大体上能猜出天气的状况。当预报今天白天“晴间多云”或“晴”，我们并不觉得稀奇，因为和我们猜测的基本一致，所得的信息量不大；如果预报是“阴有小雪”，我们就要大吃一惊，感到气候反常，这时就获得了很大的信息量，因为“阴有小雪”的先验概率很小。“阴有大雪”的先验概率更小，几乎是不可能出现

的现象，因此如果一旦出现“阴有大雪”的气象预报，我们将会万分惊讶，这时将会获得更大的信息量。由此可知，某一事物状态出现的概率越小，其不确定性越大；一旦这种状态出现，就将获得很大的信息量。

这些例子告诉我们：某一事物状态的不确定性的_{大小}，也就是我们事先猜测它是否发生的难易程度，与该事物状态的可能数目及其概率特性有关。因此，某一事物状态具有信息量的大小，与该事物可能出现的不同状态数目以及各状态出现的概率大小有关。可见，信息是可测度的。

我们把某事物各种可能出现的不同状态，即所有可能选择的消息的集合，称为**样本空间**；每个可能选择的_{消息}是这个样本空间的一个元素。对于离散消息的集合，概率测度就是对每一个可能选择的_{消息}指定一个概率（非负的，且总和为1）。一个样本空间和它的概率测度称为一个概率空间，我们用这概率空间来描述信源。

一般概率空间用 $[X, P]$ 来表示。离散情况下， X 的样本空间可写成 $[a_1, a_2, \dots, a_q]$ ，通常又称为信源符号集。样本空间中_{选择任一元素} a_i 的概率表示为 $P_X(a_i)$ ，其角标 X 表示所考虑的概率空间是 X 。如果不会引起混淆，足码可以略去，写成 $P(a_i)$ 。

在离散情况下，信源的数学模型就是一个概率空间，即

$$[X, P] = \left[\begin{matrix} a_1, & a_2, & \dots, & a_q \\ P(a_1), & P(a_2), & \dots, & P(a_q) \end{matrix} \right]$$

其中 $P(a_i)$ 就是选择信源符号 a_i 作为消息的先验概率。在接收端，对是否选择这个消息（符号） a_i 的不确定性是与 a_i 的先验概率成反比的，即对 a_i 的不确定性可表示为先验概率 $P(a_i)$ 的倒数的某一函数。我们取该函数为对数函数（理由以后再述）。并把这样定义的不确定性称为该消息（符号） a_i 的自信息：

$$I(a_i) = \log \frac{1}{P(a_i)} \quad (1.1)$$

由于信源每个符号的先验概率不一定相同，所以它们的自信息 $I(a_i)$ 也不一定相同，取其平均，则得信源每个符号的平均信息量 $H(X)$ 为：

$$H(X) = \sum_{i=1}^q P(a_i) \log \frac{1}{P(a_i)} \quad (1.2)$$

该平均信息量的表达式与统计物理学中热熵的表达式相似。在统计物理学中，热熵是一个物理系统杂乱性（无序性）的度量，这在概念上两者也有相似之处。因而我们就借用“熵”这个词把 $H(X)$ 称为熵。有时为了区别，称为“信息熵”（关于熵的概念在以后各章我们还要详细论述）。

现在再来讨论消息的传递过程。当消息（符号） a_i 在信道里传送时，假设接收端收到的消息（符号）为 b_j ，由于信道有干扰，这个 b_j 可能与 a_i 相同，也可能与 a_i 有差异。我们把条件概率 $P(a_i|b_j)$ 称为后验概率，它是接收端收到消息（符号） b_j 后发送端发的是 a_i 的概率。那么 $\log \frac{1}{P(a_i|b_j)}$ ，就是接收端收到 b_j 后对发送端选择的符号是否是 a_i 尚存在的

不确定性。于是，收信者在收到消息（符号） b_j 后，已经消除的不确定性为：先验的不确定性减去尚存在的不确定性。我们把这个量定义为互信息，即

$$I(a_i, b_j) = \log \frac{1}{P(a_i)} - \log \frac{1}{P(a_i|b_j)} \quad (1.3)$$

可见, 互信息 $I(a_i, b_j)$ 是一种消除的不确定性的度量。

如果信道没有干扰, 信道的统计特性使 a_i 以概率 1 传送到接收端。这时, 收信者接到消息后, 尚存在的不确定性就等于零, 即 $P(a_i|b_j) = 1$, $\log \frac{1}{P(a_i|b_j)} = 0$, 不确定性全部消除。由此, 得互信息

$$I(a_i, b_j) = I(a_i) \quad (1.4)$$

以上就是从不确定性的概念和概率测度出发的信息度量。可以看出, 自信息和熵是对消息某种统计特性的一种定量描述。而收信者收到一个消息以后, 其所获得的信息量(互信息)既与信源的先验统计特性有关, 也与信道的特性有关。

这样定义的信息概念有许多优点。首先, 它是一个科学的定义, 有明确的数学模型。其次, 它与日常用语中的信息的含意并不矛盾。例如, 设某一事件 a_i 发生的概率等于 1, 既 a_i 是预料中一定会发生的必然事件, 如果事件 a_i 果然发生了, 收信者将不会得到任何信息(日常含义), 因为他早知道 a_i 必定发生, 不存在任何不确定性。根据公式(1.1), 因为 $P(a_i) = 1$, 所以得 $I(a_i) = \log \frac{1}{P(a_i)} = 0$ 即自信息等于零。反之, 如果 a_i 发生的概率很小, 即猜测它是否发生的不确定性很大, 一旦 a_i 果然发生了, 收信者就会觉得很意外和惊讶, 获得的信息量很大。根据公式(1.1), 因为 $P(a_i) \ll 1$, 也得 $I(a_i) = \log \frac{1}{P(a_i)} \gg 1$ 。

再者, 定义排除了日常用语中对信息一词某些主观上的含意。根据上述定义, 同样一个消息对任何一个收信者来说, 所得到的信息量(互信息)都是一样的。虽然, 同样一个消息对不同的人常常有不同的主观价值或主观意义。

上述信息概念的定义也有其局限性。首先, 我们已经看到, 这个定义的出发点是假定信源可以用一个从经典集合论为基础的概率模型来描述。但实际存在的某些信源要寻找一个合适的概率模型往往是非常复杂的。并且对某些信源来讲, 是否存在这样一种模型还值得探讨。

其次, 这个定义是不考虑收信者的主观特性和主观意义的。也撇开了事件本身的具体含意、重要程度和引起后果等等因素。这就与实际情况不完全一致。例如当听了同一广播或看了同一报文后, 对不同的收信者来说常引起不同的关心程度、不同的感情, 因此应认为是获得了不同的信息。

为此, 有关信息的概念及其数学模型的研究还在不断地深入。目前, 信息已被人们看作是物质世界的基本属性之一。现代哲学家和科学家认为: 物质、能量、信息是物质世界的三大支柱, 是科学历史上三个最重要的基本概念。而这三者之间有着内在的密切联系。物质运动的动力是能量, 而信息是关于物质运动的千差万别的状态的知识。只要有运动的事物, 就需要有能量, 也就会存在信息, 可见信息是普遍存在的。因此, 可以说人类在改造世界和认识世界的一切活动中无不牵涉到信息的交换和利用。此处信息的概念已远远超出原来通信的范畴。可以预言, 随着人们对信息这一概念的不断深化, 信

息论和信息科学将会揭示出客观世界和人类主观世界的更多的内在规律，从而使人们有可能创造出更多更好的信息系统——信息获取系统、信息传输系统、信息控制系统以及信息处理系统。

1.2 信息论研究的对象、目的和内容

从上节关于信息概念的讨论中，我们已经看到：信息论的研究对象是消息传输系统。由于消息中包含着信息，所以消息传输系统也是信息传输系统。简称通信系统。人们通过消息的传输和处理过程来研究信息传输和处理过程中的共同规律。

信息论最初是在狭义的通信系统基础上总结提高而发展起来的，但理论一旦从实践中抽象出来，它的意义就远远超越了原来的范围。这里所谈的通信系统是广义的，不只是指电报、电话等一类狭义的通信系统，而是指所有的信息流通系统。如生物有机体的遗传系统、神经系统等，甚至人类社会的管理系统都属于广义的通信系统。

我们把这种广义的通信系统概括成如图1.2所示的统一模型。这个模型主要分成五个部分。

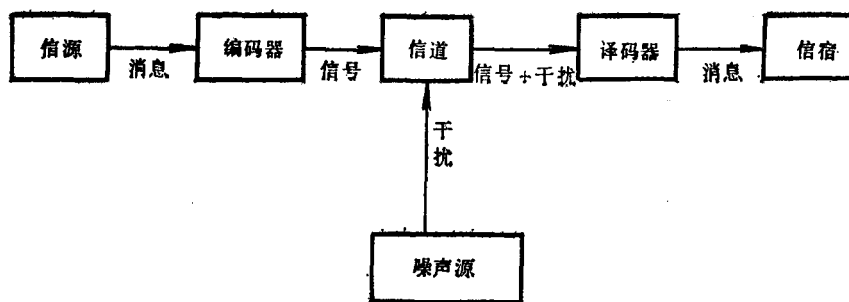


图 1.2 通信系统模型

（1）信息源（简称信源）。顾名思义，信源是产生消息和消息序列的源。如前所述，“母亲身体情况”、“各种气象状态”等客观存在是信源，人的大脑思维活动也是一种信源。信源的输出是消息，消息是具体的，但它不是信息本身。消息携带着信息，消息是信息的表达者。

（2）编码器。在许多场合，为了克服时间或空间的限制而进行通信，必须对消息进行加工处理，把消息变换成适合信道传输的物理现象。这种物理现象称为信号（如电信号、光信号、声信号等等）。所以编码是把消息变换成信号的措施，而译码就是编码的反变换。编码器输出的是适合信道传输的信号，信号携带着消息，它是消息的负荷者。

编码器可分为二种，即信源编码器和信道编码器。为了提高信息传输的有效性和可靠性，它们分别对原始消息进行适当的处理。当然，对于各种实际的通信系统，编码器还应包括换能、调制、发射等各种变换处理。

（3）信道。信道是指通信系统把载荷消息的信号从甲地传输到乙地的媒质。在狭义的通信系统中实际信道有明线、电缆、波导、光纤、无线电波传播空间等等，这些都是属于传输电磁波能量的信道。当然，对广义的通信系统来说，信道还可以是其他的传输

媒质。

信道除了传送信号以外，还有存储信号的作用。如书写通信方式就是一例。

(4)译码器。译码就是把信道输出的信号(已迭加了干扰)进行编码的反变换。一般认为这种变换是可逆的。译码器也可分成信源译码器和信道译码器。

(5)信宿。信宿是消息传送的对象，即接收消息的人或机器。

在模型中，我们在信道中引入噪声和干扰，这是一种简化的表达方式。为了方便起见，把系统其他部分产生的干扰和噪声都等效地折合成信道干扰，看成是由一个噪声源产生的，它将作用于信道中传输的信号上。

研究这样一个概括性很强的通信系统，其目的就是要找到信息传输过程的共同规律。一旦总结出这种共同规律，就可以用来指导具体通信系统的设计，使设计出来的各种通信系统具有更高的可靠性和有效性。

所谓可靠性高，就是要使信源发出的消息经过信道传输以后，尽可能准确地不失真地再现在接收端。而所谓有效性高，就是经济效果好，即用尽可能短的时间和尽可能少的设备来传送一定数量的信息。两者的结合就能使系统达到最优化。

以后我们会看到，提高可靠性和提高有效性常常会发生矛盾，这就需要统筹兼顾。例如为了兼顾有效性(考虑经济效果)，有时就不一定要求绝对准确地在接收端再现原来的消息，而是可以允许一定的误差或一定的失真，或者说允许近似地再现原来的消息。

香农(C. E. Shannon)在1948年发表的文章的序言中有这样一句话：“通信的基本问题是要在某一端准确地或近似地再现从另一端选择出来的消息”。这句话恰如其分地表达了信息论研究的目的是为了提提高通信系统的可靠性和有效性。

关于信息论研究的具体内容，是一个有争议的问题。有人认为信息论只是概率论的一个分支，这是数学家的观点。当然，这种看法是有一定根据的，因为香农信息论确实为概率论开拓了一个新的分支。但如果把信息论限制在数学范围内，这就太狭隘了。也有人认为信息论只是熵的理论，这是某些物理学家的观点。他们对“熵”特别感兴趣，熵的概念确实是香农信息论的基本概念之一，但信息论的全部内容要比熵的概念广泛得多。

目前，对信息论研究的内容一般有以下三种理解：

(1)信息论基础。主要研究信息的测度、信道容量以及信源和信道编码理论等问题。这部即是香农基本理论。

(2)一般信息论。主要也是研究信息传输和处理问题。但除了香农理论以外，还包括噪声理论、信号滤波和预测、统计检测与估计理论、调制理论以及信息处理理论等。后一部分内容是以美国科学家维纳(N. Wiener)为代表，其中最贡献的是维纳和苏联科学家哥尔莫哥洛夫(A. Колмогоров)。

虽然维纳和香农等人都是运用概率和统计数学的方法来研究准确地或近似地再现消息的问题，都是为了使消息传送和接收最优化，但他们之间却有一个重要的区别。

维纳研究的重点是在接收端。研究一个信号(消息)如果在传输过程中被某些因素(如噪声、非线性失真等)所干扰时，在接收端怎样把它恢复、再现，从干扰中提取出来。在此基础上，创立了最佳线性滤波理论(维纳滤波器)、统计检测与估计理论，噪声理论等。

而香农研究的对象则是从信源到信宿之间的全过程，是收、发端联合最优化问题，其重点是放在编码。他指出，只要在传输前后对消息进行适当的编码和译码，就能保证在干扰的存在下，最佳地传送和准确或近似地再现消息。为此发展了信息测度理论、信道容量理论和编码理论等等。

(3) 广义信息论。广义信息论不仅包括上述二方面的内容，而且包括所有与信息有关的领域，如心理学、遗传学、神经生理学、语言学、语义学甚至包括社会学中有关信息的问题。

本书遵循第二种理解，把其它所有与信息有关的领域都看成是信息论的应用。

综上所述，信息论是一门应用概率论、随机过程、数理统计和近代代数方法，来研究广义的信息传输、提取和处理系统中一般规律的工程科学；它的主要目的是提高信息系统的可靠性和有效性以便达到系统最优化；它的主要内容(或分支)包括香农理论、编码理论、维纳理论、检测和估计理论、信号设计和处理理论、调制理论和随机噪声理论等。

由于信息论研究的内容极为广泛，而各分支又有一定的相对独立性，因此本书仅论述信息论的基础理论即香农理论。

* 1.3 信息论发展简史和现状

信息论从诞生到今天，已有三十多年历史，现已成为一门独立的理论科学，回顾它的发展历史，我们可以知道理论是如何从实践中经过抽象、概括、提高而逐步形成的。

信息论是在长期的通信工程实践和理论研究的基础上发展起来的。

通信系统是人类社会的神经系统，即使在原始社会也存在着最简单的通信工具和通信系统，这方面的社会实践是悠久漫长的。

电的通信系统(电信系统)已有150年的历史了。在一百余年的发展过程中，一个很有意义的历史事实是：当物理学中的电磁理论以及后来的电子学理论一旦有某些进展，很快就会促进电信系统的创造发明或改进。这是因为通信系统对人类社会的发展，其关系实在是太密切了。日常生活、工农业生产、科学研究以及战争等等，一切都离不开消息传递和信息流动。

例如，当法拉第(M. Faraday)于1820年~1830年期间发现电磁感应的基本规律后，不久莫尔斯(F. B. Morse)就建立起电报系统(1832~1835)。1876年，贝尔(A. G. Bell)又发明了电话系统。

1864年麦克斯韦(Maxell)预言了电磁波的存在，1888年赫兹(H. Hertz)用实验证明了这一预言。接着1895年英国的马可尼(G. Marconi)和俄国的波波夫(A. C. Попов)就发明了无线电通信。本世纪初，(1907年)根据电子运动的规律，福雷斯特(L. Forest)发明了能把电磁波进行放大的电子管。之后很快出现了远距离无线电通信系统。大功率超高频电子管发明以后，电视系统就建立起来了(1925~1927)。电子在电磁场运动过程中能量相互交换的规律被人们认识后，就出现了微波电子管(最初是磁控管，后来是速调管、行波管)，接着，在三十年代末和四十年代初的二次世界大战初期，微波通信系统、微波雷达系统等就迅速发展起来。五十年代后期发明了量子放大器。六十年代初发明的激光技术，使人类进入了光纤通信的时代。

随着工程技术的发展,有关理论问题的研究也在逐步深入。

1832年莫尔斯电报系统中高效率编码方法对后来香农的编码理论是有启发的。

1885年凯尔文(L.Kelvin)曾经研究过一条电缆的极限传信率问题。

1922年卡逊(J.R.Carson)对调幅信号的频谱结构进行了研究,并建立了信号频谱概念。1924年奈奎斯特(H.Nyquist)指出,如果以一个确定的速度来传输电报信号,就需要一定的带宽。他把信息率与带宽联系起来了。

1928年哈特莱(R.V.Hartley)发展了奈奎斯特的工作,并提出把消息考虑为代码或单语的序列。在 s 个代码中选 N 个码即构成 s^N 个可能的消息。他提出“定义信息量 $H = N \log s$ ”,即定义信息量等于可能消息数的对数。其缺点是没有统计特性的概念。他的工作对后来香农的思想是有影响的。

1936年阿姆斯特朗(E.H.Armstrong)认识到在传输过程中增加带宽的办法对抑制噪声干扰肯定有好处。根据这一思想他提出了宽偏移的频率调制方法,该方法是有划时代意义的。

1939年达德利(H.Dudley)发明了声码器。当时他提出的概念是:通信所需要的带宽至少同所要传送的消息的带宽一样宽。达德利和莫尔斯都是信源编码的先驱者。

但是,一直到三十年代末,理论工作的一个主要弱点是把消息看成是一个确定性的过程。这就与许多实际情况不相符合。当时所依靠的数学工具主要是经典的付里叶分析方法,这是有局限性的。

四十年代初期,由于军事上的需要,维纳在研究防空火炮的控制问题时,提出了“平稳时间序列的外推、内插与平滑及其工程应用”的论文。他把随机过程和数理统计的观点引入通信和控制系统中来,揭示了信息传输和处理过程的统计本质。他还利用早在三十年代初他本人提出的“广义谐波分析理论”对信息系统中的随机过程进行谱分析。这就使通信系统的理论研究面貌焕然一新,引起了质的飞跃。

1948年香农在贝尔系统技术杂志上发表了两篇有关“通信的数学理论”的文章。在这两篇论文中,他用概率测度和数理统计的方法系统地讨论了通信的基本问题,得出了几个重要而带有普遍意义的结论,并由此奠定了现代信息论的基础。

香农理论的核心是:在有噪信道中只要信息传输速率低于某个值——信道容量,则就存在着一种编码方法,它可以使消息传输过程的差错概率任意小。这个结论完全出乎人们的意料,引起了科学家们的巨大兴趣。当然,从数学观点看,香农的结论是一个最优编码的存在定理。从工程观点看,尽管香农没有提出实现最优编码的具体途径,但存在性的严格证明促进了人们去寻找有效通信系统的途径。

从五十年代开始,一部分科学家(包括香农本人)又把已经得到的数学结论作了进一步的严格论证和推广。在这方面1954年范因斯坦(A. Feinstein)的论著是有很大贡献的。香农本人在五十年代发表了许多重要文章,1959年他发表了“保真度准则下的离散信源编码定理”,以后发展成为“信息率失真理论”。这一理论是频带压缩、数据压缩的理论基础,它一直到今天还保持着继续发展的势头。

与此同时,另外一部分科学家从事寻找最佳编码(纠错码)的研究工作。在近二十几年中这一工作取得了很大的进展,并已经形成一门独立的分支——纠错码理论。

整个五十年代,维纳理论也有很大进展。对于维纳——霍夫(Wiener-Hopf)方程的

研究取得了不少成果。到五十年代后期,维纳理论已推广到非平稳过程以及时变系统中。

六十年代初,卡尔曼(R. E. Kalman)和伯西(R. S. Bucy)将状态变量法引入到滤波理论中。用信号与噪声的状态空间模型代替了协方差函数,用时域的微分方程来表示滤波问题,得到了递推滤波算法,这就是卡尔曼滤波理论。线性滤波理论的这一突破,很快就成功地应用于卫星轨道测量、导弹制导和自动化等许多领域。整个六十年代主要是向应用方面发展。进入七十年代以后,卡拉思(T. Kailath)等人发展了新息(innovation)过程的理论。这一理论不仅可以用来解决白高斯型检测问题和线性最小平方估计问题,而且也可以用来解决非高斯型检测和非线性最小平方估计以及非平稳连续过程。他们还把鞅过程(Martingale Process)理论与新息理论相结合,为非线性估计开拓了前景。

在信号检测理论方面,1943年诺思(North)提出了匹配滤波器理论。1946年苏联学者卡切尼可夫(В. А. Кателъников)提出了判断错误概率为最小的理想接收机理论。整个五十年代,统计检测理论发展迅猛。米德尔顿(D. Middleton)等人用最小平均风险准则(贝叶斯准则)来处理最佳接收问题,使检测理论发展到一个新阶段,并使各种准则统一于风险理论。到五十年代末卡庞(J. Capon)提出采用非参数统计判断方法。这一方法到七十年代初开始得到了广泛的实际应用。

目前,信息论的两个方面的内容都取得更大的发展。在香农信息论方面,当前值得注意的动向是信息概念的深化;多址和多用户信道(双向信道,广播信道,多元连接型信道等)理论的发展;多重相关信源编码理论的发展;信息率失真理论的发展及其在数据压缩和图象处理中的应用等问题。这些领域都是与八十年代信息工程的前景——空间通信、计算机网络、图象电子学等密切相关的。

已经独树一帜的纠错码理论还在继续向前发展。纠错码在工程应用方面以及最优编码方法问题都有待今后继续研究。

在维纳型信息论方面,由于光纤通信即将成为现实,成象雷达以及二维图象信息处理正在迅猛发展。为此,我们对量子检测和估计理论、非参数检测和估计理论以及非线性检测与估计理论都要给予足够的重视。

最后,值得注意的一个动向是,把香农信息论的某些成果用来解决检测与估计理论中的某些问题,特别是用来解决难度很大的非线性估计问题,如用信息率失真理论作为基础来研究非线性估计问题。另一方面,用统计检测与估计理论来发展香农理论也已做了一些工作。看来今后这两个学派的互相渗透与互相促进,必定会为八十年代带来新的成果。

第二章 离散信源

从这一章开始,我们从有效而可靠地传输信息的观点出发,对组成信息传输系统的各个部分分别进行讨论。本章首先讨论信源,重点是信源的统计特性和数学模型,以及各类离散信源的信息测度——熵及其性质。这部分内容是香农信息论的基础。

2.1 信源的数学模型及分类

信源是信息的来源,但信息是较抽象的东西,所以要通过信息的表达者——消息来研究信源。我们对信源的内部结构、为什么产生和怎样产生各种不同的消息都不作研究,而只研究信源的输出,以及信源输出各种可能消息的不确定性。

正如绪论中所述,在通信系统中受信者在未收到消息以前,对信源发出什么消息是不确定的,所以可用随机变量或随机矢量来描述信源输出的消息。或者说,用概率空间来描述信源。

有的信源输出的消息常常是以一个个符号的形式出现,例如文字、字母等,而且这些符号的取值是有限的或可数的。所以,可用离散型随机变量来描述这些消息。这样的信源称为离散信源。它的数学模型就是离散型的概率空间:

$$[X, P] = \begin{matrix} X: & [a_1, a_2, \dots, a_q] \\ P(x): & [P(a_1), P(a_2), \dots, P(a_q)] \end{matrix}$$

显然, $P(a_i) (i=1, 2, \dots, q)$ 应满足

$$\sum_{i=1}^q P(a_i) = 1$$

它表示信源可能取的消息(符号)只有 q 个: $a_1, a_2, \dots, a_q$, 而且每次必定取其中一个。

有的信源输出的消息的取值是连续的,即可能出现的消息数是不可数的无限值。可用连续型随机变量来描述这些消息。那么这种信源称为连续信源,其数学模型为连续型的概率空间:

$$[X, p] = \begin{matrix} X: & [(a, b)] \text{ 或 } [R] \\ p(x): & [p(x)] \end{matrix}$$

并满足

$$\int_a^b p(x) dx = 1 \text{ 或 } \int_R p(x) dx = 1$$

其中 R 表示实数集 $(-\infty, \infty)$, 而 $p(x)$ 是随机变量 X 的概率密度函数。

上述信源是最简单的情况,因信源只输出一个消息(符号),所以可用一维随机变量来描述。